CRV 2008: Fifth Canadian Conference on Computerand Robot Vision, Windsor, ON, Canada, May 2008

Conference participation

Fihl, Preben

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Technical Report CVMT-08-02 ISSN 1601-3646, December 2008

CRV 2008: Fifth Canadian Conference on Computer and Robot Vision, Windsor, ON, Canada, May 2008

P. Fihl

Laboratory of Computer Vision and Media Technology

Aalborg University, Denmark

Email: pfa@cvmt.dk

Summary: *This technical report will cover the participation in the fifth Canadian Conference on Computer and Robot Vision in May 2008. The report will give a concise description of the topics presented at the conference, focusing on the work related to the HERMES project and human motion and action recognition. Our contribution to the conference will also be described and discussed.*

1. General conference information

The Canadian Conference on Computer and Robot Vision (CRV) was held in Winsor, Ontario, Canada on May 28-30 2008. The University of Windsor was hosting the conference. The conference was mainly a national conference and the majority of the participants were Canadian.

The conference was a part of the Intelligent Systems Collaborative which covers the conferences: Artificial Intelligence 2008 (AI), Graphics Interface 2008 (GI), Intelligent Systems 2008 (IS), SMARTLinkages 2008 (SMART), and Computer and Robot Vision 2008 (CRV). The Intelligent Systems Collaborative had a rather strong industrial focus which was evident from the plenary sessions and the technology showcase that was taking place during breaks throughout the three conference days. Especially the stand from Point Grey Research caught interest with impressive camera demonstrations.

2. Talks of Intelligent Systems Collaborative

The plenary sessions covered industrial views on challenges within the area of Information and Communication Technology. The talks were interesting but not directly related to human motion and action recognition.

Peter Carbone, Vice President at Nortel, presented how the Canadian company Nortel is moving away from just providing transmission of telecommunication signals and towards providing end-user services. The introduction of IP-technology and the ideas from the structure of the internet was presented as the main facilitator for this change and the globalization in telecommunications made such a change necessary.

Joachim G. Taiber, Head of Information Technology Research Office at BMW Group, presented how BMW is establishing research offices to develop efficient systems for the hugely increasing amount of information that is being generated in both car manufacturing and in the everyday use of cars in the near future. The research offices both deal with management of data (e.g. database design and server-park organization) and presentation of data (e.g. in-car display of maps and email).

Rick Whittaker, Vice President at Sustainable Development Technology Canada, presented an ongoing program to support and develop Canadian companies that work within the area of information and communication technology, especially “clean tech” companies, e.g. a company that does automatic monitoring of ship traffic and thereby enables tracking of oil leaking ships.

The rest of the conference talks were running as parallel tracks, i.e. one track for each of AI, GI, IS, SMART, and CRV. CRV was the most relevant at all times with the exception of an invited talk at GI.

Dr. Patricia Jones, Acting Division Chief, Human Factors Research and Technology Division from NASA, gave an invited talk on advanced visualization and interfaces at NASA. The talk gave an overview on the organizational structure of the NASA offices working on visualization and interfaces and it showed some examples of the work. An interesting point was an extensive use of computer animated film clips. NASA uses

animated films at many stages of their work and Dr. Jones pointed to this as a very important communication tool. The films are used internally to makes sure that project members have a common conception of the project, they are used in fund raising both towards politicians and other sponsors to convey the ideas of a project, and they are used for the public in general as an educational tool and for general information. Using video clips to present scientific research is an effective approach and it deserves more attention in many research projects. Another interesting point was a currently running project that investigates how astronauts respond to color monitors in the space shuttles compared to the green screens in use at the moment. The project illustrates how conservative NASA needs to be when sending people into space.

3. Research from CRV

CRV featured seven oral sessions, two poster sessions, and three invited talks. The oral sessions were: segmentation, stereo vision, early vision, 3D vision/object recognition, robot control, motion tracking and activity recognition, and robot vision.

3.1 CRV in general

The presentations were in general informative and interesting. However, the majority of the presented work seemed to be either master projects or newly started research projects. This meant that the presented work often had quite limited extends and that especially the tests were limited.

An interesting aspect of the conference was the variety in sensor types. Color cameras, stereo cameras and laser range scanners are often used in computer and robot vision but also work on ultra sound images [2], light detection and ranging data (LIDAR) [4], and especially infra red (IR) images were presented at the conference [3][5][6]. Combinations of color cameras with IR cameras [7], mirrors (for directional blurring) [8], and fisheye lenses (for wide field of view stereo) [1] were also presented. The different sensors make vision methods applicable to new problem areas and the combination of different kinds of sensors often increase the precision or quality of the results compared to a single type of sensor. The number of multi sensor systems seems to be increasing based on the papers presented at CRV. The ongoing development of still better sensors and the fact that e.g. IR cameras are beginning to become consumer products will ensure that the increase in multi sensor systems will continue.

3.2 Invited talks at CRV

The three invited talks were: “Seeing in Stereo” by Professor Steven W. Zucker, Yale University, “Intelligent Implies Embodied, Autonomous and Situated” by Professor James L. Crowley, INRIA Grenoble, and “Underwater Robots and Vision” by Professor Greg Dudek, McGill University.

The first talk presented an approach to depth estimation from stereo images where polynomial functions were fitted to image edges. A number of results and examples were shown and a comparison was made to the traditional approach of matching image

patches. The presented approach seemed to outperform the traditional approach based both on the underlying assumptions and results that were presented.

The second talk addressed the questions of which properties a system must posses to be denoted *intelligent*. The title of the talk indicated an answer, namely that an intelligent system must be embodied, autonomous and situated, but the talk itself and the vivid discussion at the end of the talk revealed that the title was only a suggestion to a very much open question. The required properties of an intelligent system seemed to be an important question in the artificial intelligence community. In the talk some concepts of intelligent systems was exemplified by a very general system running at INRIA Grenoble with an intelligent office as the main application.

The last talk presented work on human-robot communications under water. A set of small autonomous underwater robots was developed at McGill University and the talk focused on the development a set of markers and marker-movements that enables a diver to communicate with these robots. The underwater environment was a big challenge to a computer vision system, however, the vision system was better than alternatives like sound or physical contact according to the presented research.

3.3 Research related to HERMES and the Ph.D. project

Several papers were related to the HERMES project and the Ph.D. project. Naturally, papers in the session “motion tracking and activity recognition” were relevant to the project but also a couple of other papers were relevant.

Detection of hand-to-ball events is presented in [10]. The paper presents a method to detect events such as catch, release, push, and hit ball. A ball, the hand, and the lower arm of a person are tracked in a video sequence. A simple setup and a simple tracking of the ball and the hand are adopted from other works. The focus of the paper is the modeling of the trajectories with fifth order polynomials and the use of gravitational models that enable the event detection. The tests are limited but the method seems to require good tracking and rather clear movements to work well. It could however be interesting to investigate how these principles extend to detection of events like punching, shaking hands, pushing, and blocking a punch.

[11] presents an interesting method for 3D human motion tracking. A low dimensional mapping is learnt between the human pose and image data by applying a dynamic probabilistic latent semantic analysis (PLSA). The human pose originates from either computer generated human motion or from motion capture data. The corresponding image data is in the first case rendered by computer graphics software and is in the second case a real image. A particle filter is applied in the low dimensional space to track human motion in input sequences. The method is computationally efficient compared to similar approaches and it is claimed to utilize the temporal correspondence in a better way. Multiple viewpoints can be included in the training data which enables the method to estimate an unknown viewpoint in an input sequence. The paper refers to a number of related papers (in a good related work section) and assumes that the reader has knowledge of the methods used in these papers which makes the paper hard to read for the non-expert. The results seem convincing, however, the test on real data is only evaluated visually, i.e. does the estimated pose *look like* the true pose. The authors seem

to point to the efficient processing as the most important property of the method. Applying a method like this may be reasonable to the problem of gait type classification but it does not seem reasonable to apply such a method to a general pose estimation problem since it would require too much training data.

Tracking of people is also investigated in [12]. The paper describes a detection and tracking algorithm based on corner detection. The method finds temporal correspondence between all detected corner points which gives a number of candidate trajectories. These trajectories are then classified as either background trajectories (noise) or foreground trajectories (persons). The classification is done based on three criteria. The first criterion removes stable and random trajectories. The second removes backward motion (defined in a scene specific manner). The third criterion removes trajectories that originate from a background model (acquired through a training phase with an empty scene). The tracker is simple but it works well when the motion that needs to be tracked is rather simple and when the accuracy of the trajectories is not critical. The three classification criteria are defined in a very scene specific manner but it may be possible to define similar criteria for other scenes. However, other tracking approaches, like the ones available in the HERMES project, are by far more accurate and more general.

The motion tracking and activity recognition session also featured a paper on simple hand and face detection and tracking by skin color detection which extended to a simple recognition of medication intake. The methods of the paper were very basic and their relevance to the HERMES and Ph.D. projects was therefore very limited. Another paper of that session was on object tracking in the presence of occlusion. The occlusion detection seemed quite effective but relied heavily on rigid objects which also made this paper of limited relevance to the HERMES and Ph.D. projects.

An interesting paper on detection and tracking of multiple objects was presented at the first poster session [13]. The method detects objects (people, vehicles, and bags) by extraction of a combined color and texture feature. The color-texture feature is extracted for small image blocks (8x8 pixels) and motion detection is done by analyzing the change in the feature vector over a small number of consecutive frames. Detected objects are represented by attributed relational graphs where the nodes represent sub-regions of the objects with similar color-texture features and the vertices represent spatial relations between the sub-regions. Tracking now becomes a problem of graph matching and by doing inexact graph matching the method is able to track objects through occlusions and deformations. The test results are very limited but the examples shown from the PETS dataset look very encouraging and it could be interesting to see publications on this approach at a later stage.

A region-based background subtraction method is presented in [9]. The method constructs a background model based on color histograms and texture information of rectangular regions. The background model is calculated at different scales in a hierarchical manner. At the finest level of the hierarchy the background is also modeled by a mixture of Gaussians. In the subtraction phase motion is first detected in each rectangle and only the rectangles that contain significant motion will be processed further, i.e. iteratively subdivided into new rectangles. At each level the foreground can be segmented by use of the color histogram and the texture information and at the finest level the mixture of Gaussians is applied to refine the segmentation result. The

hierarchical structure of the method makes it possible to get segmentation at different scales in an elegant way and the ideas of the region-based segmentation seem to be effective. However, the test shows that the method only performs slightly better than just using a mixture of Gaussian.

3.4 Own research at CRV

Our contribution to CRV was the paper *Invariant Classification of Gait Types* [14] which was a part of the motion tracking and activity recognition session. The paper presents a method of classifying human gait based on silhouette comparison. A database of artificially generated silhouettes is created representing the three main types of gait, i.e. walking, jogging, and running. Silhouettes generated from different camera angles are included in the database to make the method invariant to camera viewpoint and to changing directions of movement. The extraction of silhouettes is done using the Codebook method and silhouettes are represented in a scale- and translation-invariant manner by using shape contexts and tangent orientations. Input silhouettes are matched to the database using the Hungarian method. A classifier is defined based on the dissimilarity between the input silhouettes and the gait actions of the database. The overall recognition rate is 88.2% on a large and diverse test set. The recognition rate is better than that achieved by other approaches applied to similar data.

The comments that were made after the presentation concerned two aspects. The first was the amount of variability represented in the silhouette database. This issue has been raised before in reviews and in general discussions about the work so the presentation did explain this aspect in some detail. Apparently, the presentation did not fully succeed in answering potential questions. From the work with the silhouette database we have found that overall characteristics of gait can actually be modeled with just one prototype execution. This is because the individual variability in the execution of gait is somewhat subtle, and since we do not rely on appearance but only silhouettes the computer graphics training data actually contains enough information. The results presented in the paper also support this conclusion. The data used for testing comes from four different datasets and is completely decoupled from the training data, and still we achieve results that are comparable to other state of the art methods. The second question addressed the processing time. At the time of the presentation the system was implemented partly as an old program coded in C for background subtraction and partly as a prototype running in Matlab with only little attention paid to the processing time. This resulted in a processing time of a few seconds per frame not including the background subtraction (which was running at 5-10 fps). This was briefly discussed at the conference. The system has subsequently been reimplemented which has resulted in a real time gait analysis system running at approximately 15 fps.

References

[1] S. Gehrig, L. Krüger*: 6D Vision Goes Fisheye for Intersection Assistance*. CRV 2008, On., Canada.

[2] M. Zouqi, J. Samarabandu: *Prostate Segmentation from 2-D Ultrasound Images Using Graph Cuts and Domain Knowledge*. CRV 2008, On., Canada.

[3] Q. Wu, P. Boulanger, W. F. Bischof: *Robust Real-Time Bi-Layer Video Segmentation Using Infrared Video*. CRV 2008, On., Canada.

[4] W. Lu, J. J. Little, A. Sheffer, H. Fu: *Deforestation: Extracting 3D Bare-Earth Surface from Airborne LiDAR Data*. CRV 2008, On., Canada.

[5] M. Akhloufi, A. Bendada: *Thermal Faceprint: A New Thermal Face Signature Extraction for Infrared Face Recognition*. CRV 2008, On., Canada.

[6] M. Akhloufi, A. Bendada: *Hand and Wrist Physiological Extraction for Near Infrared Biometrics*. CRV 2008, On., Canada.

[7] F. Morin, A. Torabi, G. Bilodeau: *Automatic Registration of Color and Infrared Videos Using Trajectories Obtained From a Multiple Object Tracking Algorithm*. CRV 2008, On., Canada.

[8] B. Reskó, P. Baranyi*: Opto-Mechanical Oriented Edge Filtering*. CRV 2008, On., Canada.

[9] P. D. Z. Varcheie, M. Sills-Lavoie, G. Bilodeau: *An Efficient Region-Based Background Subtraction Technique*. CRV 2008, On., Canada.

[10] N. Miller, R. Mann: *Detecting Hand-Ball Events in Video*. CRV 2008, On., Canada.

[11] K. Moon, V. Pavlovic: *3D Human Motion Tracking Using Dynamic Probabilistic Latent Semantic Analysis*. CRV 2008, On., Canada.

[12] N. M. E. Nabbout, J. Zelek, D. Clausi: *Automatically Detecting and Tracking People Walking Through a Transparent Door with Vision*. CRV 2008, On., Canada.

[13] W. Huang, Q. M. J. Wu: *Detection and Tracking of Multiple Moving Objects in Real-World Scenarios using Attributed Relational Graph*. CRV 2008, On., Canada.

[14] P. Fihl, T. B. Moeslund: *Invariant Classification of Gait Types*. CRV 2008, On., Canada.